Goal for Lustre HA

• Detect failures and provide architecture to deal with any level of Failure

• Provide continued access to data with no to minimal I/O interrupts for applications

• Multiple redundant components is the basis for Lustre® Storage Availability
Data Protection Layer

- At minimum design solution around RAID 6 or write data based on 8+2 (8 data blocks and 2 parity blocks)
- Need to plan for rare cases of multiple drive failures within a single storage enclosure
  - Use of Hot Spares
  - Balance Front-end OST Performance with back-end disk rebuilds
- Adjusting rebuilds on the fly can protect against rare cases
Individual HA Domains

- Two Lustre Servers in a HA pair having access to all the OSTs
  - Reliable redundant Heartbeats
    - Two different networks or the use of passive back-plane within a storage enclosure and external network
  - Dual-Path Drives and SAS Paths
  - Separate Fabric Switches for Server in the Pair
HA Event Detection

- HA services detect and failover resources when loss of data from clients occur
  - Fabric Failure
  - SAS Interruption to Disks
  - Heartbeat Compromised
  - Software Interrupt
- HW Failure of any kind
- Rules and timeout values are critical in HA services to ensure the discovery of a failure is handled within short period of time
HA Failover

• When a partner detects an issue with a problem node, want to avoid dual-mount, split brain, or an event causing two nodes to fail in a HA Domain

• Defining the rules is critical and the use of STONITH

• In an event a healthy nodes questions it’s partner, Shoot The Other Node In The Head to avoid the problem node causing issues to healthy partner

• When partner node powers off unhealthy node, resources will failover automatically

• On failure, once Lustre Targets are mounted, resources can still be unavailable to client due to recovery time
  • Depends on number of Clients, and amount of data to replay when Failover occurred
Failback

• When a HA event occurs, Failback should be manual
  • Automatic failback can be problematic if the node is still unhealthy and can cause more problems
• Failback is done manually for various reasons
  • Admins can do maintenance to ensure the problem causing the failover/STONITH is fixed
• Avoid further outage due to Lustre Recovery Time on failback
• Depending on architecture, suggest having module solution so on failure it will effect performance on a small percentage of overall solution
Fabric Connectivity

- If connecting all storage to a single Leaf Switch or Core Switch, this is the SPOF.
- Ensure the solution can survive a complete fabric switch failure in addition to:
  - HCA Failure
  - HCA Driver issue
- Cable Issue
- If using Leaf-Core module, multiple fabric links from Leaf to each core is highly recommended to provide redundant cable connectivity.
UPS and Power

- UPS is critical to protect against unexpected power glitches or interrupts to power.
- Without UPS, any power interruption can compromise data integrity or disk drive MTBF.
- UPS can provide enough power to survive short power interrupts or allow for graceful shutdown of Lustre.
- Multiple Power grids to feed the redundant rack PDUs provide redundancy if one power feed is compromised.
Integration and Testing in Factory

• Hardware integration
  • Integrating the storage solution designed from the ground up into a single chassis proves to be highly available compared to non-integrated solutions

• Building, configuring and testing the entire storage solution in factory improves increased reliability and reduces on-site integration

• Repeatable performance on benchmarks for proven throughput
Thank You

John _Fragalla@xyratex.com